Eye-Gaze Tracking in Near-Eye Head-Mounted Displays

Alexander Plopski
plopski@is.naist.jp
Arno in Wolde Luebke
arno-w@is.naist.jp
Wataru Kashima
kashima.wataru.ks3@is.naist.jp
Takafumi Taketomi
takafumi-t@is.naist.jp
Christian Sandor
sandor@is.naist.jp
Hirokazu Kato
kato@is.naist.jp
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan

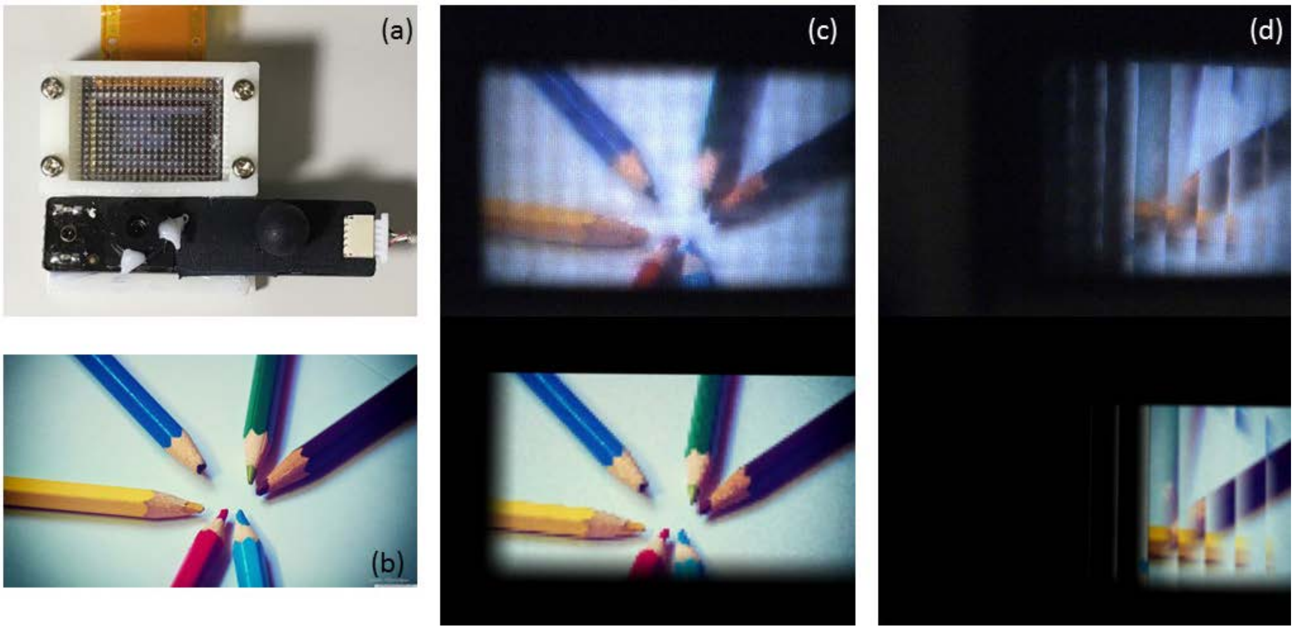

Figure 1: The effects of eye movement on the image observed on the microlens HMD. (a) Our microlens display equipped with an eye-tracking camera. (b) The image we aim to display on the HMD (taken from `www.ultrahdwallpapers.net`**). The view of the display taken with a camera (top row) and our simulation (bottom row). (c) Both methods display the correct image when the camera is located near the position used for the generation of the HMD content. (d) When the real and virtual cameras are shifted without adapting the rendering the view becomes distorted with artefacts.**

ABSTRACT
Existing near-eye microlens display designs do not account for variations in the user's eye pose. This can lead to incorrectly rendered content thus degrading the overall experience. In this paper we describe the design of a microlens display equipped with an eye-gaze tracking camera. We describe how to calibrate the HMD so that the camera can be used for accurate 6DOF eye-pose estimation, present an approach to reduce the blur of presented images and present a method to simulate the user's perception. Initial experiment results suggest that our approach provides a more accurate representation than images taken with a camera. In the future an evaluation of the system will provide important information for design of near-eye wide-field-of view HMDs.

OzCHI '16, November 29 – December 2 2016, Launceston, TAS, Australia

http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx.

Author Keywords
Head-Mounted Display, Calibration, Eye-Pose Estimation, Microlens Display

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces Graphical User Interfaces.

INTRODUCTION
Since the announcement of the Occulus Rift consumer interest in Head-Mounted Displays (HMDs) has reached a new high with various models being announced by a number of well-known manufacturers. Although the design of these devices is more user friendly than the first HMD envisioned by Sutherland et al. (Sutherland, 1968), it still entails a bulky casing in front of the user's eyes. This casing incorporates the display(s), magnifying lenses, and some mechanical elements to adjust their position. Additionally, a relatively large spacing distance between the lenses and the display is necessary to place the virtual image within the user's accommodation range. Otherwise, the image becomes blurred and unobservable.

Researchers have proposed a variety of manufacturing approaches that could lead to thin-form, large field-of-view HMDs; e.g., microlens arrays (Lanman and Luebke, 2014), and pinhole projectors (Maimone et al., 2014). Although these displays could help elevate the form problems of current HMDs, their usability has not been verified beyond controlled experiments. In particular, the authors assume that the position of the eye is known and verify the applicability of their designs with cameras whose parameters are set to be similar to the anatomy of the human eye.

We show the significance of eye-gaze tracking (EGT) for near-eye microlens displays by taking images of the display from different positions with a camera (Figure 1, top row). Given a target scene (Figure 1(b)) we generate content on the display that would create the desired view. When the camera's position coincides with the assumed position the recorded image resembles the target view (Figure 1(c)). However, when the camera's position shifts (Figure 1(d)) aberrations become more dominant and the image quality degrades.

Our ultimate goal is to evaluate the impact errors in eye-pose estimation have on the perceived image and to determine the accuracy required to display consistent images for various eye poses in near-eye, microlens HMD setups. In this paper we make three contributions towards it:

- We describe the design and calibration of an eye-gaze tracking capable microlens HMD based on work of (Lanman and Luebke, 2014).
- We describe a modified approach to generation of content displayed on the HMD screen that reduces the blur effects and accounts for pixels being seen through multiple lenses.
- We describe how user-perspective views can be generated in a simulation. Our rendered correctly represent the user's views and are comparable to images taken with a camera.

EYE-POSE ESTIMATION

EGT has been applied in HMDs for user attention evaluation, rendering manipulation (Murphy and Duchowski, 2001), as well as interaction (Tanriverdi and Jacob, 2000). Recently, a number of commercial solutions have also included EGT-capabilities (e.g., FOVE, SMI). However, most of these solutions do not focus on the estimation of the eye-pose, but the estimation of the eye-gaze on the display. For our work it is necessary to recover not only the estimated gaze position, but also the eye-pose relative to the HMD display.

Eye-pose estimation refers to the estimation of the eye-position and orientation relative to the eye-tracking camera. This involves the recovery of eye-features from the camera image and the reconstruction of the eye-pose up to the optical axis. The offset between the optical axis and the visual axis, the actual gaze direction, can be determined through a one-time calibration step (Hansen and Ji, 2010).

The eye-pose can be estimated with passive methods that recover the eye-pose from the camera image taken under natural illumination. Commonly, the iris contour is detected as an ellipsoid in the camera image, followed by a spatial reconstruction given known anatomical parameters (Nitschke et al., 2011). The reconstructed iris contour is known up to an ambiguity that must be resolved through prior knowledge, constraints, or assumptions.

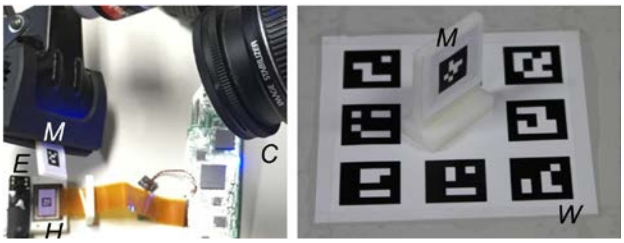

Figure 2: Calibration of the eye-tracking camera E relative to the HMD with an external camera C, marker setups M and W, and marker H displayed on the HMD screen.

Active methods make use of active scene illumination and prior scene calibration to achieve a higher degree of accuracy than passive methods. The most robust and widely applied method is pupil-center corneal-reflections (PCCR) approach (Guestrin and Eizenman, 2006) that reconstructs the position of the cornea; the eye surface that covers the iris; from the detected reflection of at least 2 known infra-red LEDs. The orientation of the eye is recovered from the pupil-contour that is reconstructed given the estimated cornea position. (Plopski et al., 2015) have adapted PCCR to use with passive estimation, however the accuracy of their method is worse than that of PCCR.

In this work we aim to use PCCR to estimate the eye-pose as it has proven to be the most accurate solution and recovers the 6DOF pose of the eye.

HMD CALIBRATION

We extend the design of (Lanman et al., 2014) with a PupilLabs eye-gaze tracking camera (Figure 1a). This camera is equipped with two infra-red LEDs and an IR filter. Thus, it fulfills the requirements for 6DOF geometrical eye-pose estimation with PCCR.

To align the estimated eye-pose with the HMD display we perform a one-time calibration step where we estimate the transformation from the camera to the HMD coordinate system. Hereby, we use two fiducial markers M_1 and M_2 that are attached to both sides of a small plate *M*, a set of markers *W* rigidly attached to the table, and a refocusable camera *C* (Fig. 2). We place *M* next to *W*, so that both sides of the plate can be recorded, while also capturing multiple markers of *W*. We record multiple images of the arrangement with *C* and determine the transformation

$$T_{M_2}^{M_1} = T_{M_2}^{W} T_{W}^{M_1}. \quad (1)$$

In a second step, we place *M* so that the pose of M_1 can be estimated by the eye-tracking camera *E*. Additionally, we display a marker *H* on the HMD, whose pose coincides with the HMD screen *D*. The size of the marker *H* is known from the number of pixels and the pixel pitch. We estimate the pose of M_2 and *H* with *C*, and compute the transformation from the eye-tracking camera to the HMD as

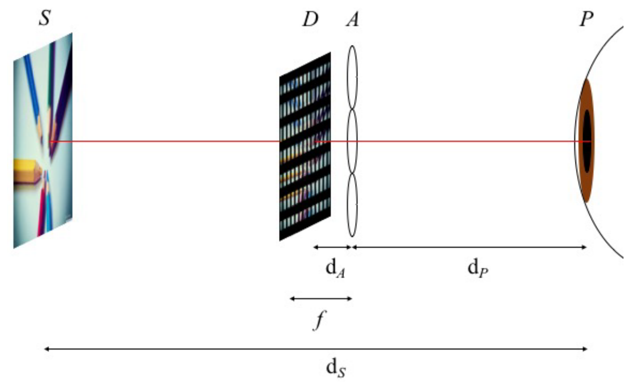

Figure 3: The schematic view of the microlens HMD. The user's pupil is located d_P away from the microlens array *P* and has a radius of r_P. The microlens array *A* is placed d_A away from the HMD screen *D* and the resulting image *S* appears to be d_S away from the user.

$$T_H^E = T_H^C T_C^{M_2} T_{M_2}^{M_1} T_{M_1}^E. \quad (2)$$

Given this transformation, the estimated eye-pose can be transformed into the coordinate system of the HMD-screen and used as input of the rendering pipeline.

IMAGE GENERATION

The original rendering method by (Lanman, 2014) does not account for pixels on the HMD screen being seen through multiple lenses of the microlens array. In this case the image seen by the user no longer corresponds to the intended view. We, therefore, propose a modified rendering algorithm that is explained in the following.

W.l.o.g. assume the setup shown in Fig. 3. The user is looking at the display so that the center of the pupil *P* with a radius r_P, the center of the microlens array *A*, and the center of the display *D* all lie on a single axis. We assume that all microlenses have the same pitch w_A and focal length f, and all pixels have the same size w_D. The microlens array is located a distance $0<d_A<f$ from *D* and the pupil is located d_P away from *A*. The screen appears to be

$$d_S = d_P + \frac{f d_A}{(d_A+f)} \quad (3)$$

in front of the user. As $d_A<f$, each pixel p_D on the display gets magnified by a factor

$$M = \frac{(d_S - d_P)}{d_A}. \quad (4)$$

To determine the content to be rendered on p_D, we check if it can be seen through multiple lenses at the same time. As shown in Fig. 4, if p_D is seen through a lens l_0 it corresponds to pixel $p_{S,0}$, given as

$$p_{S,0} = l_o + (d_S - d_P)\frac{(l_0 - p_D)}{d_A}. \quad (5)$$

The pixel $p_{S,0}$ can be seen through l_0, if any ray from $p_{S,0}$ through l_0 hits the pupil.

Now, consider a second lens l_1 ($l_1 \neq l_0$). When seen through l_1, p_D corresponds to the pixel

$$p_{S,1} = l_1 + (d_S - d_P)\frac{(l_1 - p_D)}{d_A}. \quad (6)$$

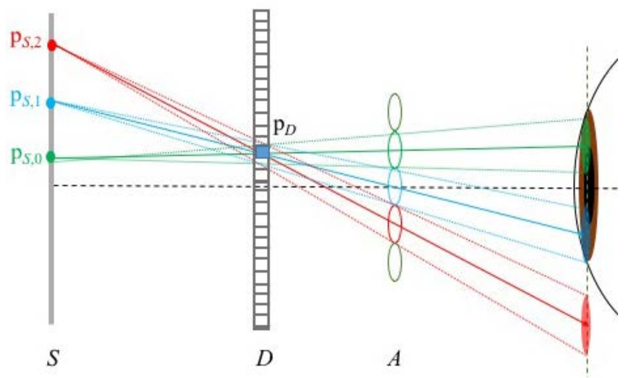

Figure 4: Computation of the color of a pixel p_D on the HMD screen depending on the pupil size. When seen through different lenses, pixel p_D corresponds to pixels $p_{S,1}$, $p_{S,2}$, and $p_{S,3}$ on the virtual screen. The pixel $p_{S,2}$ cannot be seen through the lens l_2. However, the light from pixels $p_{S,0}$ and $p_{S,1}$ enters the pupil through the corresponding lenses. Therefore, the color of pixel p_D is not uniquely defined.

If any ray passing from $p_{S,1}$ through l_1 hits the pupil, the color of p_D would be ambiguous. Therefore, we do not display any content on it. We show the image generated with the original algorithm and our method in Fig. 5.

USER VIEW SIMULATION

Top evaluate the impact of the eye-tracking and our rendering method it is necessary to recover images seen be the user. As it is not possible to see exactly what the user is observing, previously the view was approximated with images taken by a camera. However, in this setup it is difficult to apply small displacements to evaluate the error tolerance of the system. In this section we describe our approach to create controlled simulated views that correspond to the user's perspective.

The user's view is often assumed to coincide with a pinhole camera, e.g., for Optical See-Through Head-Mounted Display Calibration (Tuceryan, 2002). In practice, each point on the retina is illuminated not by a single ray, but by a number of rays that are refracted through the eye lens and contribute to the observed image. We simulate this with a camera located at the pupil's center, an aperture that equals the pupil size and is focused at the screen distance. The imaging plane image plane of the cameras is fixed at 20 mm, which roughly corresponds to the anatomy of a human eye.

We use distributed ray tracing (Cook, 1984) for image generation. Hereby, for each point on the image, we uniformly sample the lens and trace rays to the corresponding point on the object plane. Upon hitting a microlens we refract the ray according to its focal lens and compute the intersection with the display. The displayed value corresponds to the radiance incident along the path. The final irradiance at the image point is given as an average of the radiance along all incident paths. We show the results of our simulation for each camera position in Figure 1 (bottom row).

Interestingly, even though the images taken with the camera appear sharper, we found that the results of our simulation more closely resemble the images observed on the HMD with the naked eye. This supports our assumption that correct reproduction of the human

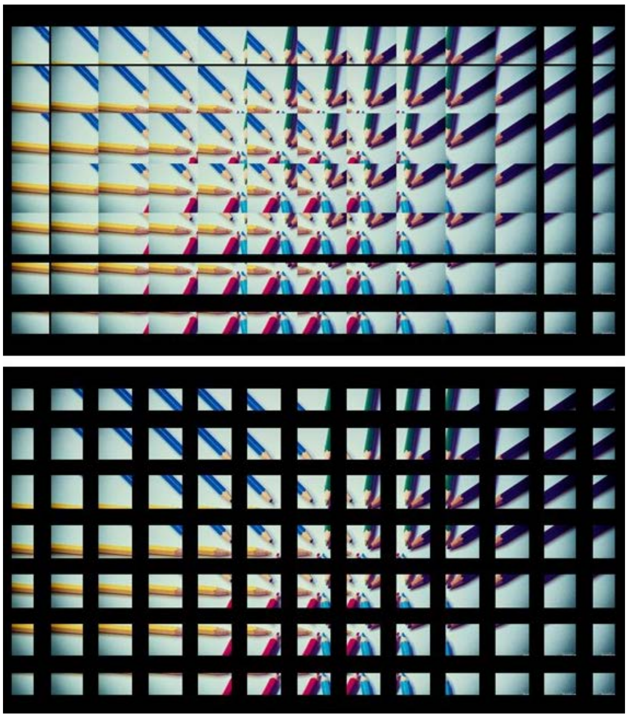

Figure 5: (Top) Screen rendering generated with the method of (Lanman, 2014) and (Bottom) our approach.

perception is necessary to evaluate the impact of rendering errors on users.

CONCLUSION

In this paper we introduced the design of an EGT capable microlens HMD. By tracking the user's eye, it will be possible to present correct images even if the HMD moves on the head, or the user moves the eye to observe a different region. We have developed a method that accounts for overlapping views of the content by the user. Using this method, we will be able to improve the rendering results and present sharper images that more closely match the desired view. To evaluate the impact of eye-tracking and our rendering approach we have developed a simulation method that approximates the image perceived by the user when looking at our HMD screen.

In the future we aim to quantify acceptable EGT and calibration error margins; e.g., errors in the estimation of the eye pose, the pupil size, or the camera pose relative to the HMD; and how these impact the image formation. Additionally, we want to perform a user study to verify the results of the simulation and the applicability of our design.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP15H01700.

REFERENCES

Akşit, K., Kautz J., and Luebke D. Slim Near-Eye Display using Pinhole Aperture Arrays. Applied optics 54, 11 (2015), 3422-3427.

Cook, R. L., Porter T., and Carpenter L. Distributed Ray Tracing. ACM SIGGRAPH Computer Graphics 18, 3 (1984), ACM, 137–145.

Guestrin, E. D. and Eizenman M. General Theory of Remote Gaze Estimation using the Pupil Center and Corneal Reflections. IEEE Transactions on Biomedical Engineering 53, 6 (2006), 1124–1133.

Hansen D. and Ji Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 3 (2010), 478–500.

Lanman, D, and Luebke D. Near-Eye Light Field Displays. ACM Transactions on Graphics 32, 6 (2013).

Maimone A., Lanman D., Rathinavel K., Keller K., Luebke D., and Fuchs H. Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses using Defocused Point Light Sources. ACM Transaction on Graphics 33, 4 (2014).

Murphy, H. and Duchowski, A. T. Gaze-Contingent Level of Detail Rendering. EuroGraphics 2001 (2001).

Nitschke C., Nakazawa A., and Takemura H. Display-Camera Calibration using Eye Reflections and Geometry Constraints. Computer Vision and Image Understanding 115, 6 (2011), 835–853.

Plopski A., Nitschke C., Kiyokawa K., Schmalstieg D., and Takemura H. Hybrid Eye Tracking: Combining Iris Contour and Corneal Imaging. Proc. ICAT-EGVE 2015, Eurographics (2015), 183–190.

Sutherland, I. E. A Head-Mounted Three Dimensional Display. Proc. AFIPS 1968, ACM Press (1968), pp. 757-764.

Tanriverdi V. and, Jacob R. J. K.: Interacting with Eye Movements in Virtual Environments. Proc. CHI 2000, ACM Press (2000), 265–272.

M. Tuceryan, Y. Genc, and N. Navab. Single-Point Active Alignment Method (SPAAM) for Optical See-Through HMD Calibration for Augmented Reality. Presence 11, 3 (2002), 259–276.